

Claims

(1) A speech recognition device configured to include a computer, the speech recognition device comprising:

a storage area for storing a feature quantity acquired from a speech signal for each frame;

storing portions for storing acoustic model data and language model data, respectively;

an echo adaptation model generating portion for generating echo speech model data from a speech signal acquired prior to a speech signal to be processed at the current time point and using the echo speech model data to generate adapted acoustic model data; and

recognition processing means for utilizing said feature quantity, said adapted acoustic model data and said

language model data to provide a speech recognition result of the speech signal.

- (2) The speech recognition device according to claim 1; wherein said adapted acoustic model generating means comprises:

a model data area transforming portion for transforming cepstrum acoustic model data into linear spectrum acoustic model data; and

an echo prediction coefficient calculating portion for adding said echo speech model data to said linear spectrum acoustic model data to generate an echo prediction coefficient giving the maximum likelihood.

- (3) The speech recognition device according to claim 2, further comprising an adding portion for generating echo speech model data; wherein said adding portion adds the cepstrum acoustic model data of said acoustic model and cepstrum acoustic model data of an intra-frame transfer

characteristic to generate a speech model affected by intra-frame echo influence.

(4) The speech recognition device according to claim 3; wherein said adding portion inputs said generated speech model affected by intra-frame echo influence into said model data area transforming portion and causes said model data area transforming portion to generate linear spectrum acoustic model data of said speech model affected by intra-frame echo influence.

(5) The speech recognition device according to claim 4; wherein said echo prediction coefficient calculating portion uses at least one phoneme acquired from an inputted speech signal and said echo speech model data to maximize likelihood of the echo prediction coefficient based on linear spectrum speech model data.

(6) The speech recognition device according to claim 5; performing speech recognition using a hidden Markov model.

(7) A speech recognition method for causing a speech recognition device configured to include a computer to perform speech recognition; the method causing the speech recognition device to execute steps of:

storing in a storage area a feature quantity acquired from a speech signal for each frame;

reading from said storing portion a speech signal acquired prior to a speech signal to be processed at the current time point to generate echo speech model data;

processing a speech model stored in a storing portion to generate adapted acoustic speech model data and store it in a storage area; and

processing said feature quantity, said adapted acoustic model data, and language model data stored in a storing portion to generate a speech recognition result of the

speech signal.

- (8) The speech recognition method according to claim 7; wherein the step of generating said adapted acoustic model data comprises steps of:

an adding portion calculating the sum of said read speech signal and an intra-frame transfer characteristic value; and

a model data area transforming portion to read said sum calculated by said adding portion to transform cepstrum acoustic model data into linear spectrum acoustic model data.

- (9) The speech recognition method according to claim 8, further comprising a step of:

causing an adding portion to read and add said linear spectrum acoustic model data and said echo speech model

data to generate an echo prediction coefficient giving the maximum likelihood.

(10) The speech recognition method according to claim 9; wherein the step of transformation into said linear spectrum acoustic model data comprises a step of causing said adding portion to add the cepstrum acoustic model data of said acoustic model and cepstrum acoustic model data of an intra-frame transfer characteristic to generate a speech model affected by intra-frame echo influence.

(11) The speech recognition device according to claim 10: wherein the step of generating said echo prediction coefficient comprises a step of determining the echo prediction coefficient so that the maximum likelihood is given to at least one phoneme for which the sum value of the linear spectrum echo model data of said speech model affected by intra-frame echo influence and said echo speech model data, which has been generated by said adding portion and stored.

(12) A computer-readable program for causing a computer to execute the speech recognition method comprising the steps of:

storing in a storage area a feature quantity acquired from a speech signal for each frame;

reading from said storing portion a speech signal acquired prior to a speech signal to be processed at the current time point to generate echo speech model data;

processing a speech model stored in a storing portion to generate adapted acoustic speech model data and store it in a storage area; and

processing said feature quantity, said adapted acoustic model data, and language model data stored in a storing portion to generate a speech recognition result of the speech signal.

(13) A storage medium storing a computer-readable program for causing a computer to execute a speech recognition method, said method comprising the steps of:

storing in a storage area a feature quantity acquired from a speech signal for each frame;

reading from said storing portion a speech signal acquired prior to a speech signal to be processed at the current time point to generate echo speech model data;

processing a speech model stored in a storing portion to generate adapted acoustic speech model data and store it in a storage area; and

processing said feature quantity, said adapted acoustic model data, and language model data stored in a storing portion to generate a speech recognition result of the speech signal.